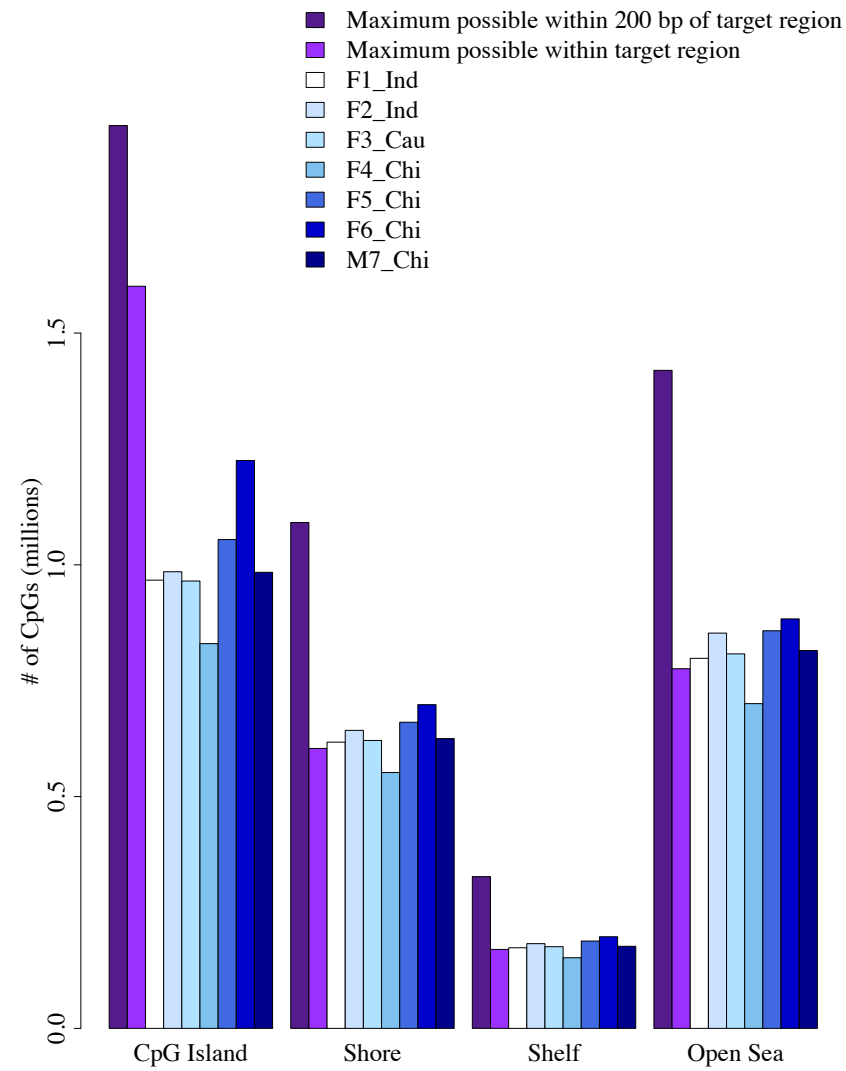


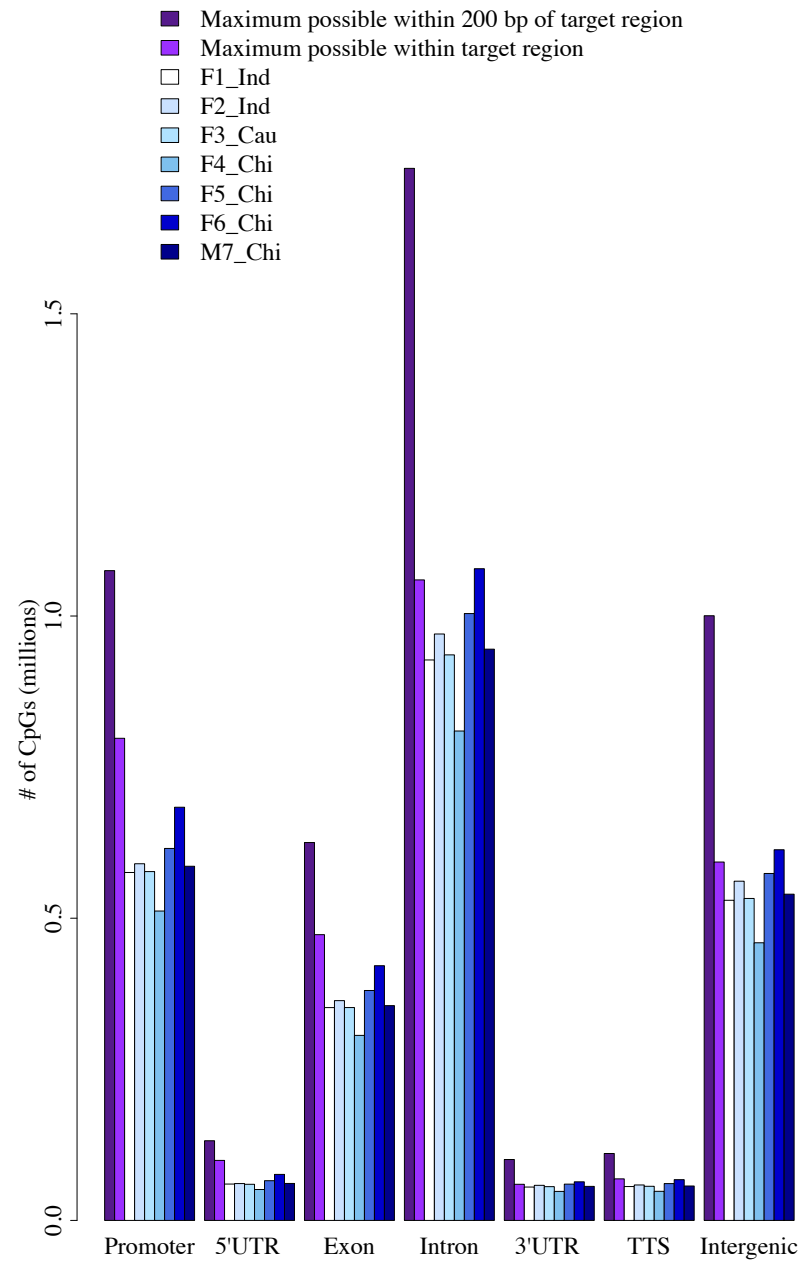
Comparison of Methyl Capture Seq vs. Infinium 450K array for methylome analysis in clinical samples

Supplementary Figures and Tables

Supplementary Figure 1: CpG content distribution of CpGs.

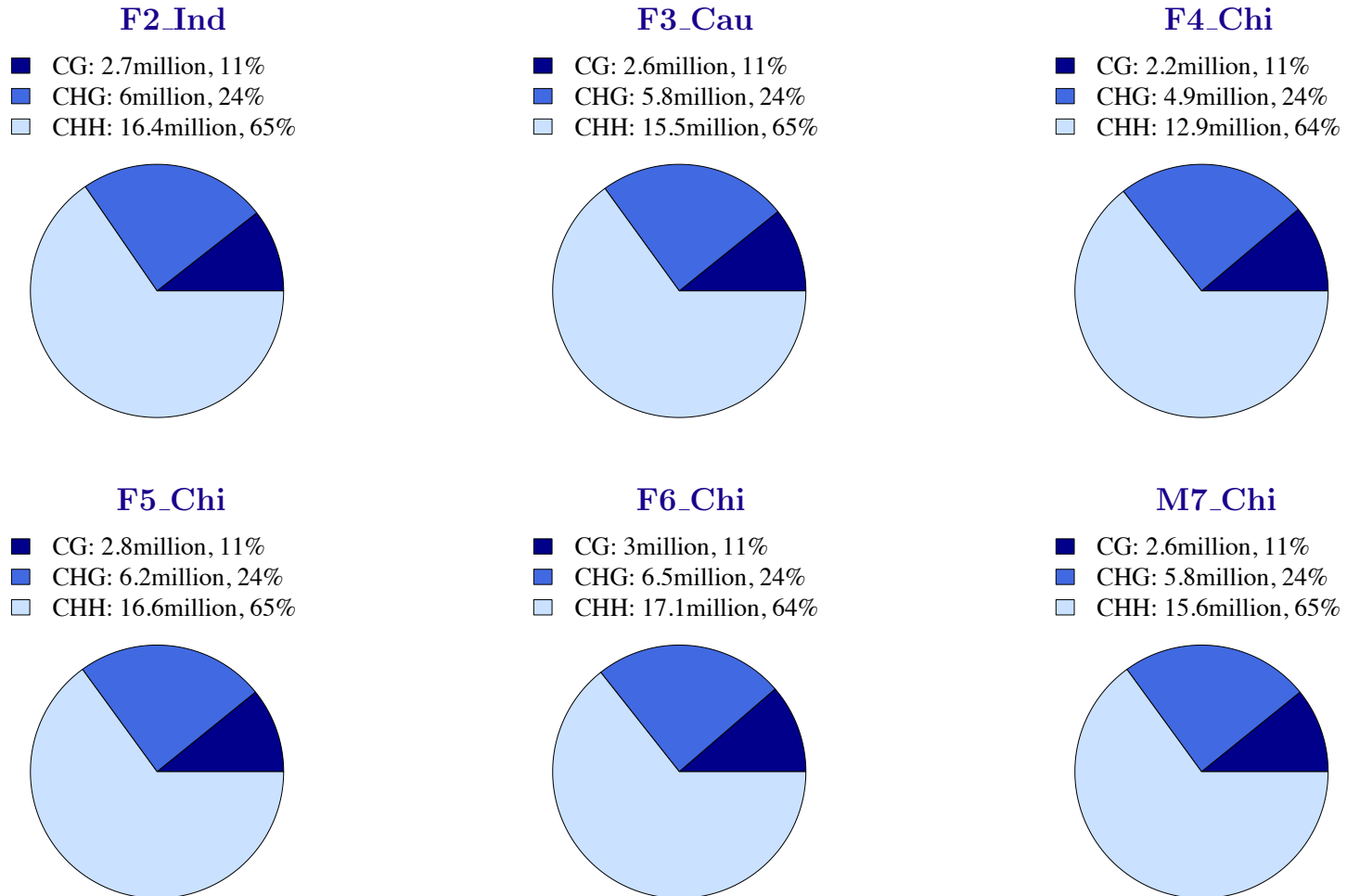


Supplementary Figure 2: Functional genomic distribution of CpGs.



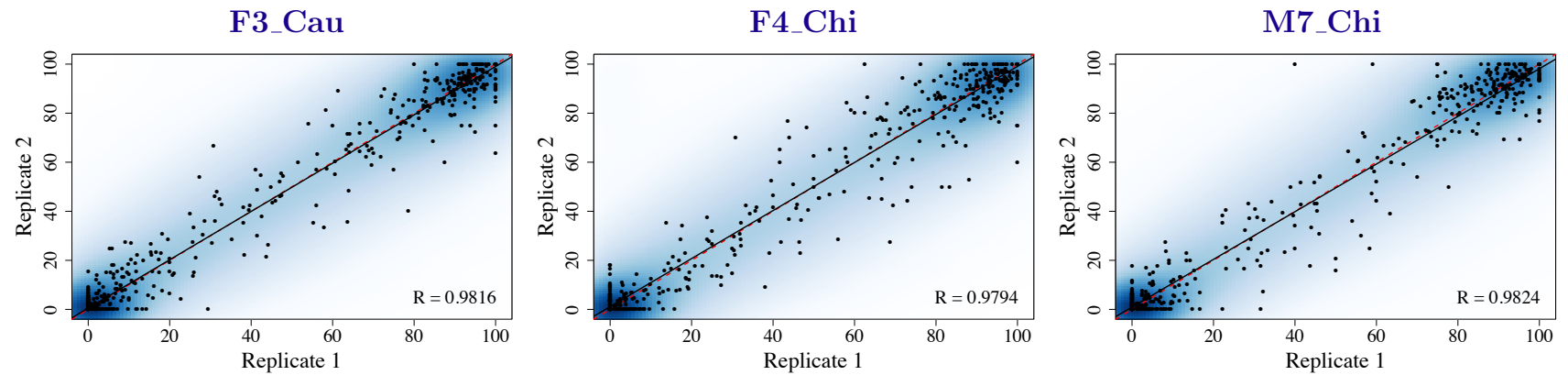
Supplementary Figure 3: Methylation sites detected within CpG/CHG/CHH context.

Methylation context (CpG/CHG/CHH) distribution ($\geq 10X$) for



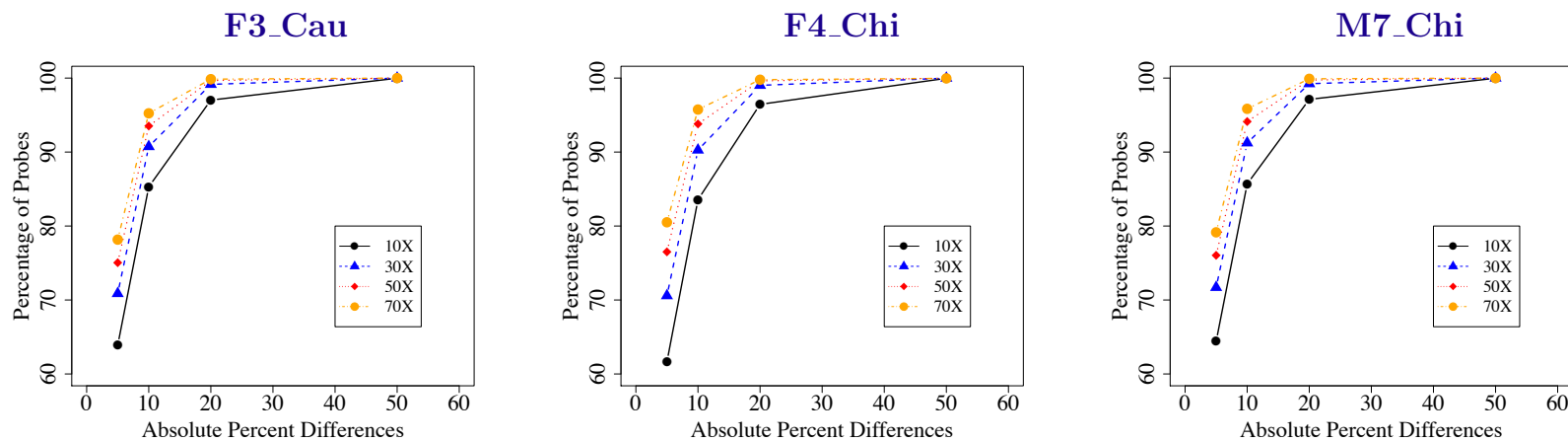
Supplementary Figure 4: Pearson correlation and scatterplot of methylation values from MC Seq for replicate 1 (horizontal axis) and replicate 2 (vertical axis).

Scatterplots of MC Seq methylation values from replicates ($\geq 10X$) for

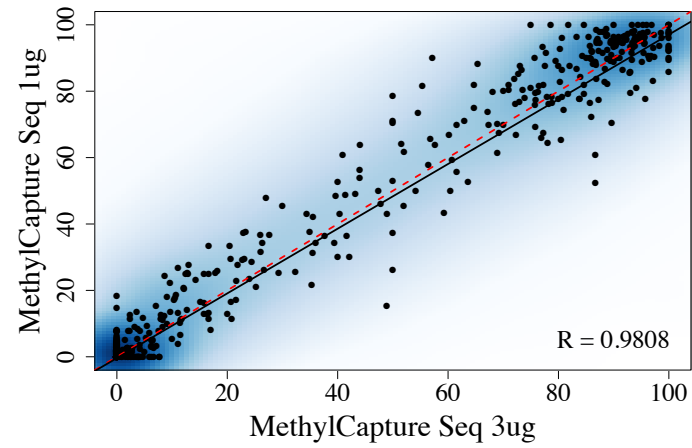


Supplementary Figure 5: Cumulative percentage of probes (vertical axis) vs. absolute difference in methylation between replicates (horizontal axis), at $\geq 10X$ (solid line), $\geq 30X$ (dashed line), $\geq 50X$ (dotted line) and $\geq 70X$ (dotted-dashed line) reads coverage.

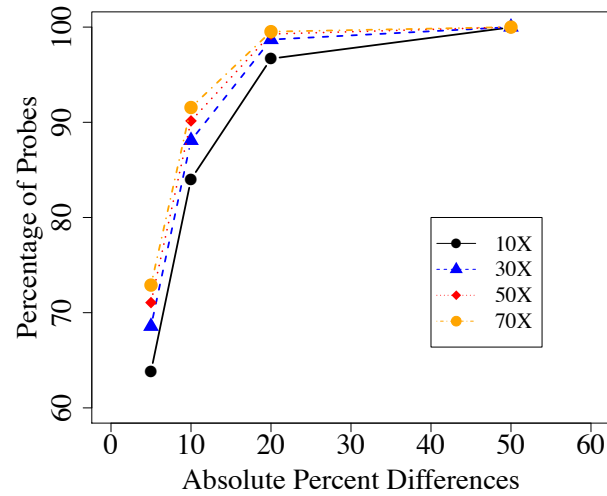
% of probes vs. absolute differences in MC Seq methylation values from replicates ($\geq 10X$) for



Supplementary Figure 6: Pearson correlation and scatterplot of methylation values from MC Seq at 3 μg (horizontal axis) and 1 μg (vertical axis) for one sample.

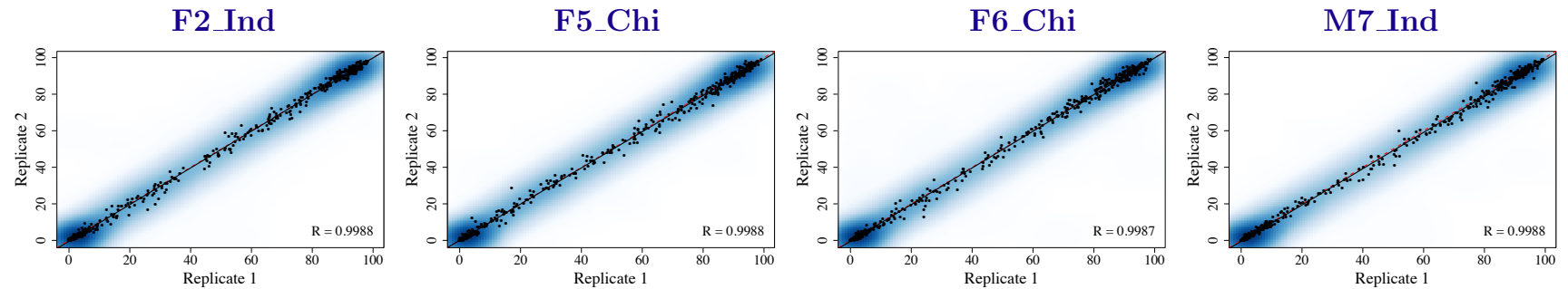


Supplementary Figure 7: Cumulative percentage of probes (vertical axis) vs. absolute difference in methylation between 3 μg and 1 μg (horizontal axis), at $\geq 10\text{X}$ (solid line), $\geq 30\text{X}$ (dashed line), $\geq 50\text{X}$ (dotted line) and $\geq 70\text{X}$ (dotted-dashed line) reads coverage.



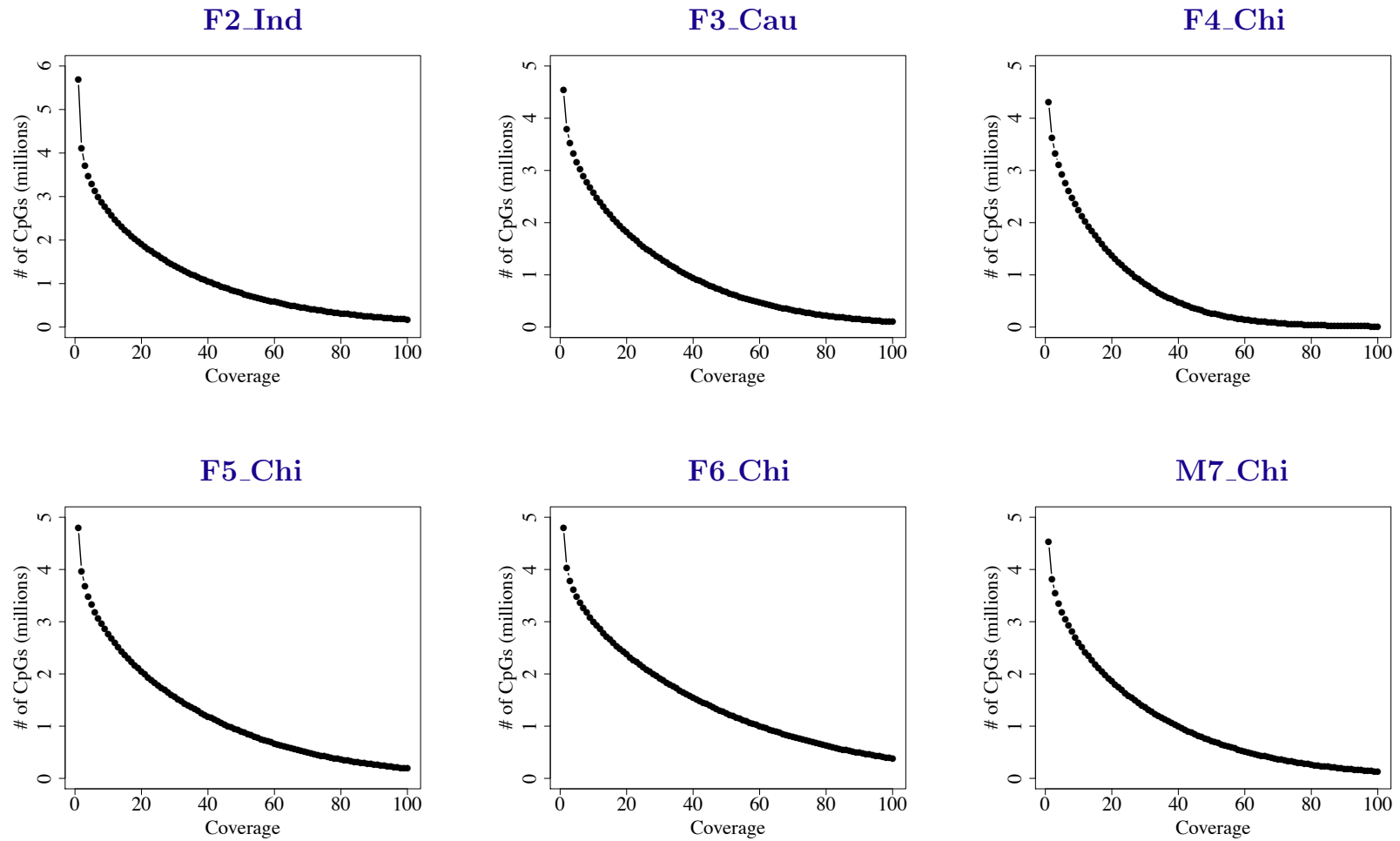
Supplementary Figure 8: Pearson correlation and scatterplot of methylation values from Infinium 450K for replicate 1 (horizontal axis) and replicate 2 (vertical axis).

Scatterplots of Infinium 450K methylation values from technical replicates for



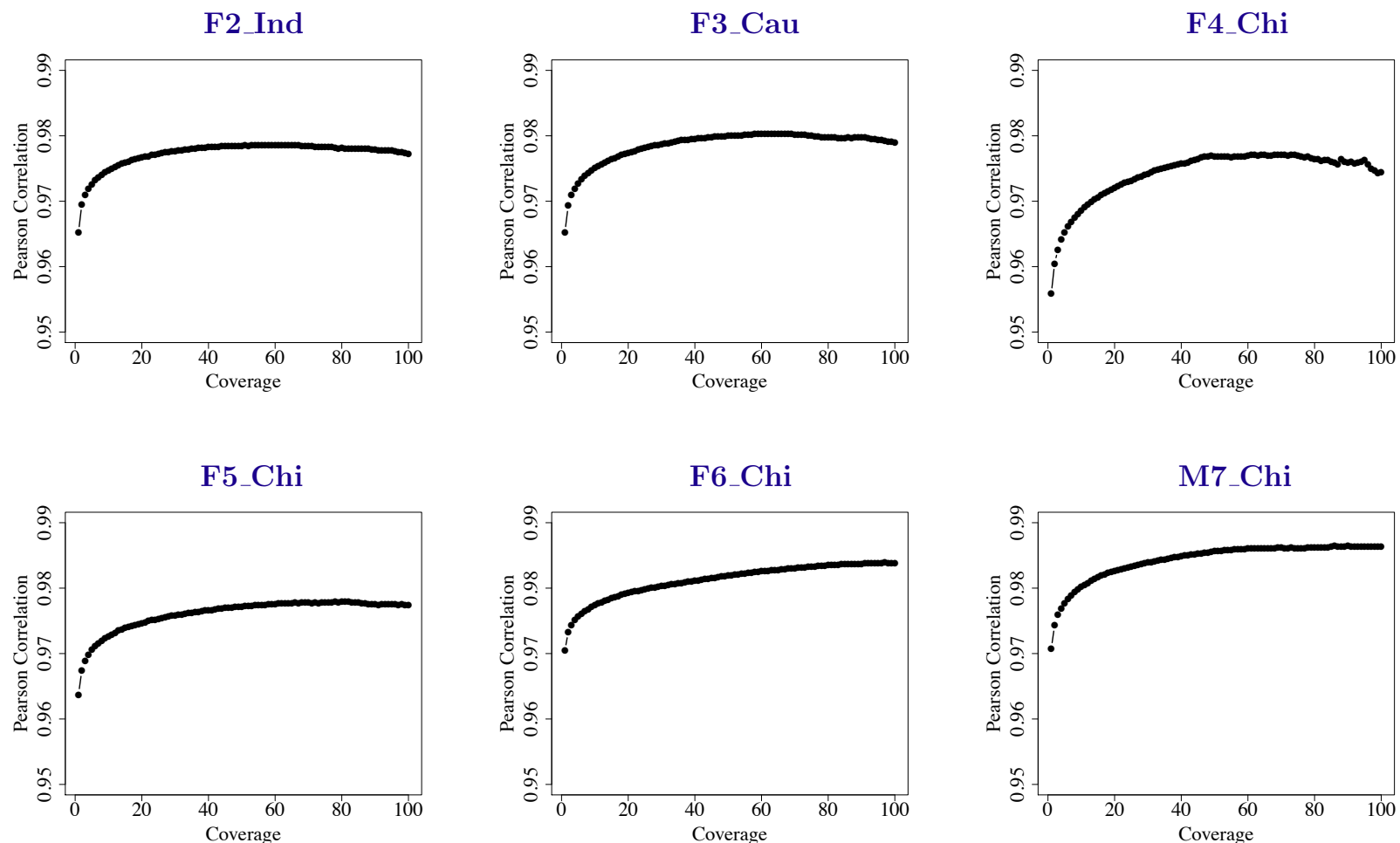
Supplementary Figure 9: Observed number of CpGs (vertical axis) from MC Seq at different MC Seq reads coverage (horizontal axis).

Number of CpGs detected by MC Seq (in millions) vs. reads coverage for



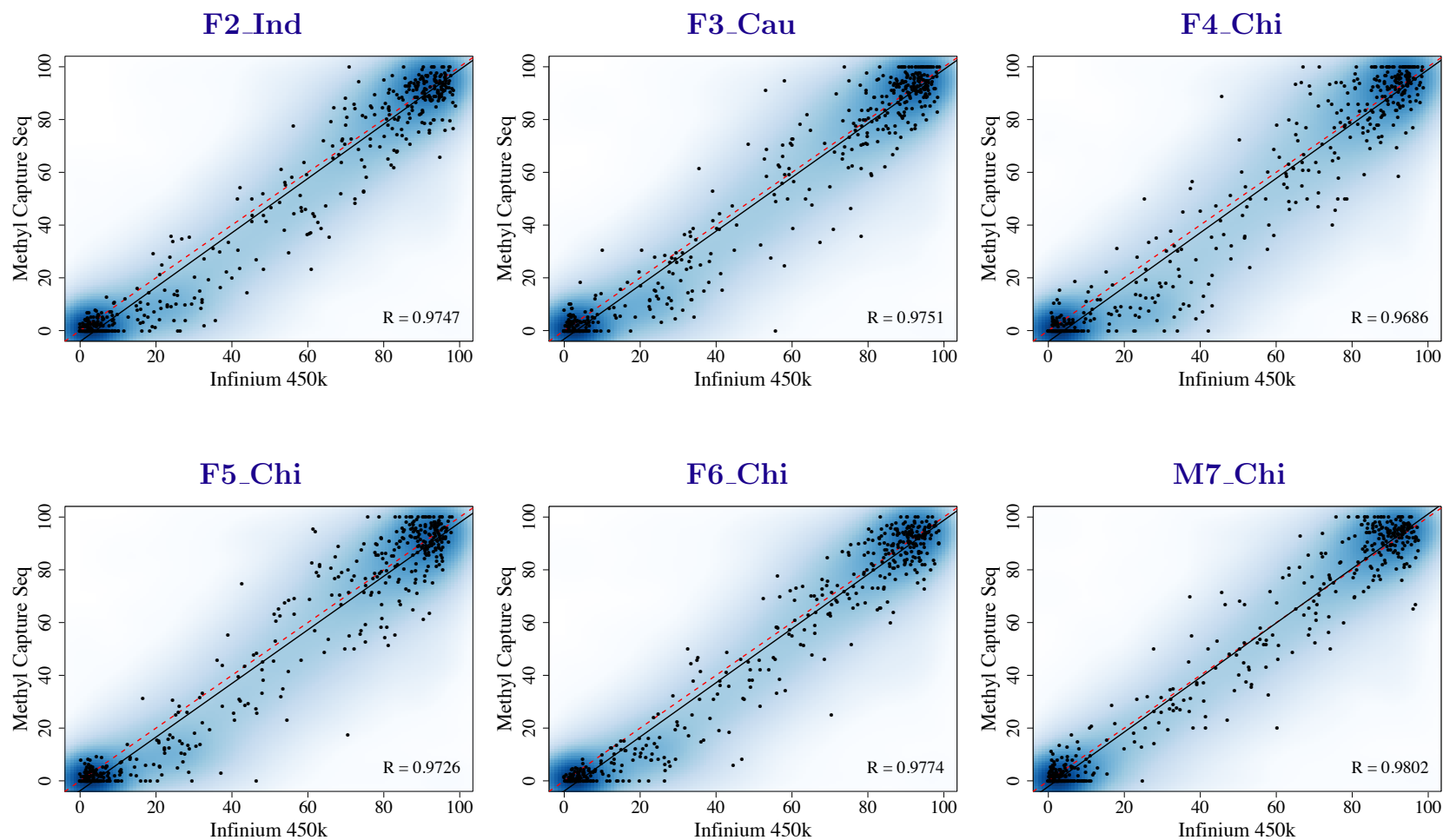
Supplementary Figure 10: Pearson correlation (vertical axis) between methylation values from MC Seq and Infinium 450K at the same CpG sites, at different MC Seq reads coverage (horizontal axis).

Pearson correlation of methylation values (between MC Seq & Infinium 450K) vs. reads coverage for



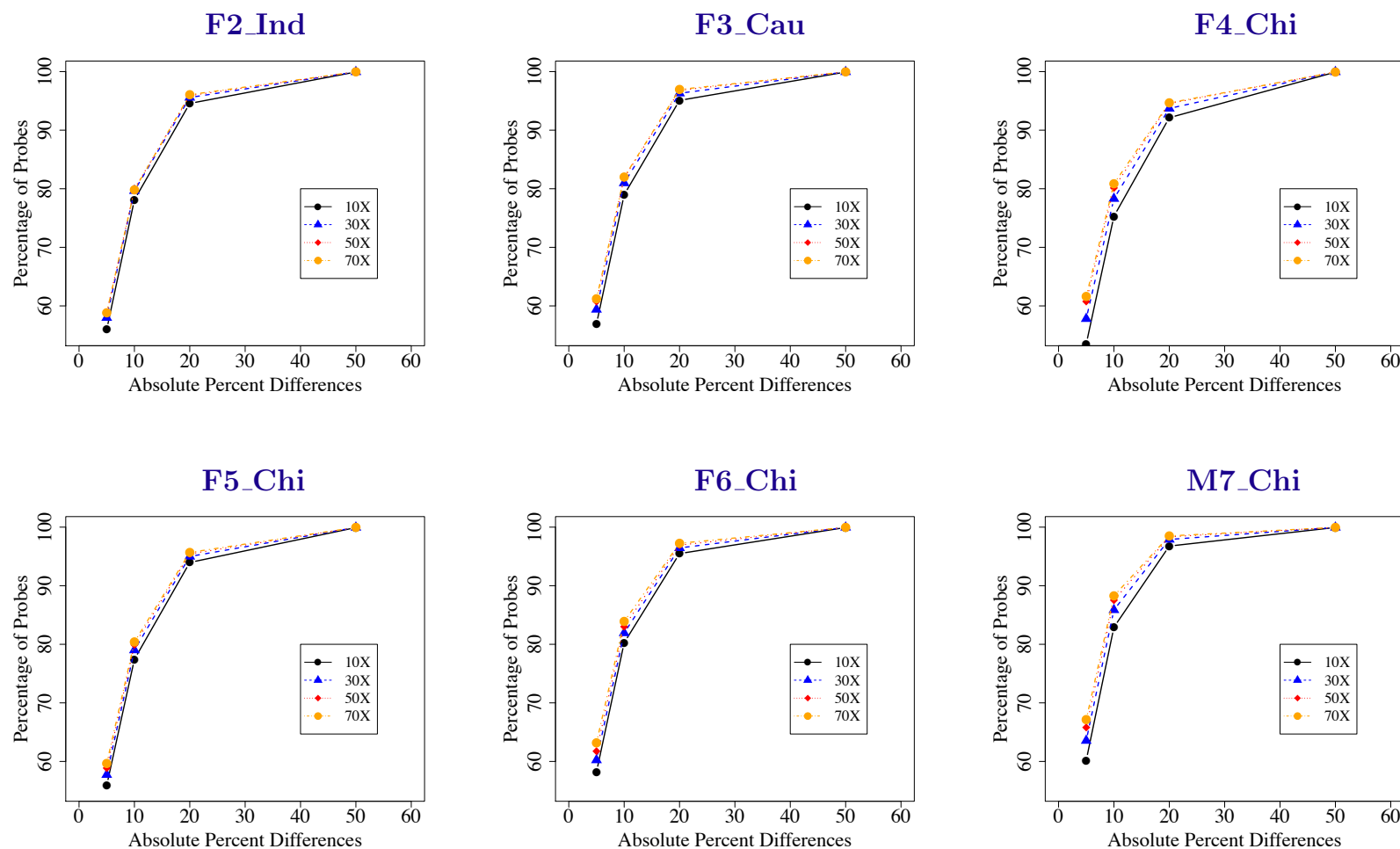
Supplementary Figure 11: Scatterplot of methylation values from MC Seq ($\geq 10X$, vertical axis) and Infinium 450K (horizontal axis) at the same CpG sites.

Scatterplots of methylation values from MC Seq ($\geq 10X$) & Infinium 450K for



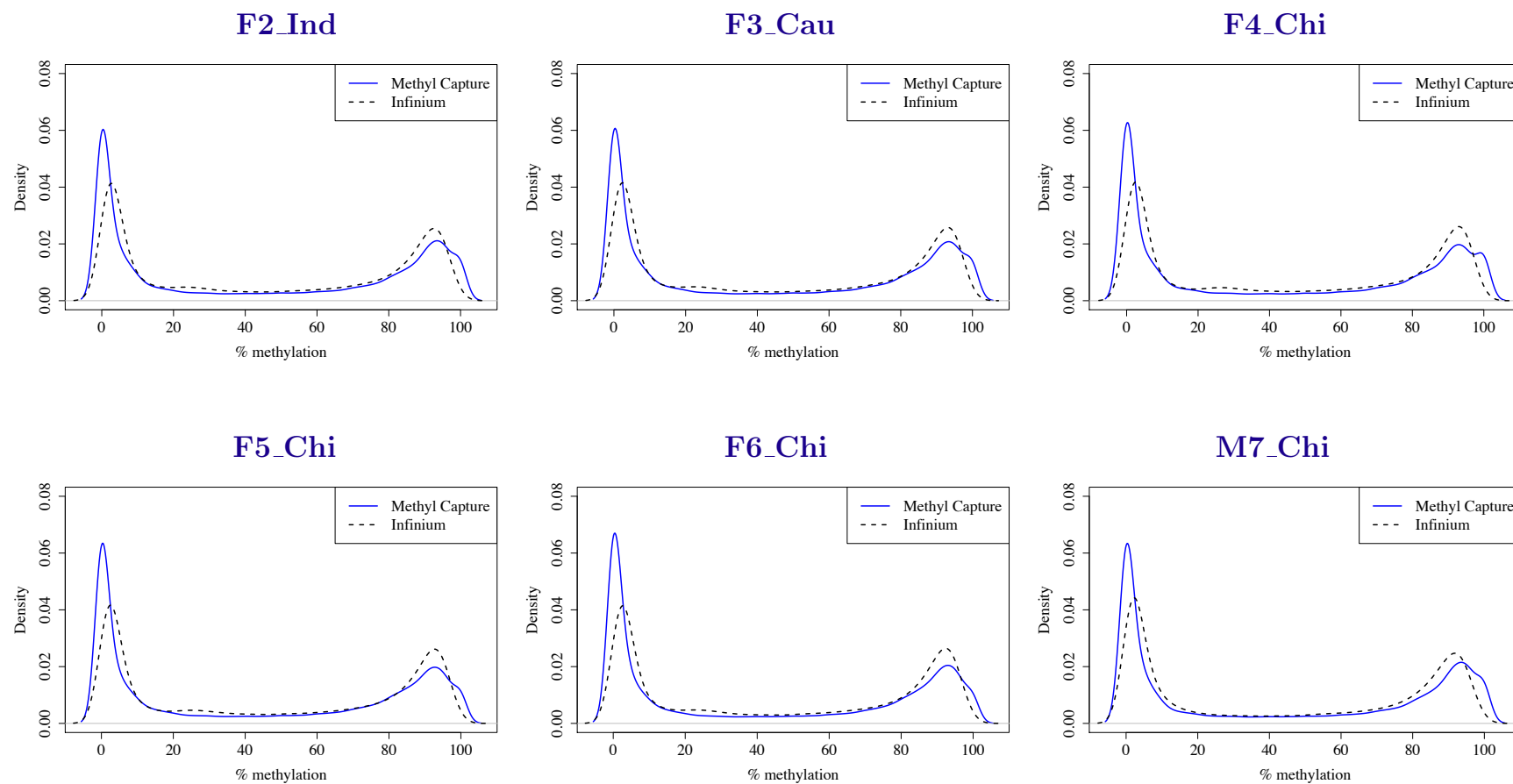
Supplementary Figure 12: Cumulative percentage of probes (vertical axis) vs. absolute difference in methylation between MC Seq and Infinium 450K (horizontal axis), at $\geq 10X$ (solid line), $\geq 30X$ (dashed line), $\geq 50X$ (dotted line) and $\geq 70X$ (dotted-dashed line) reads coverage.

Absolute differences in methylation values (between MC Seq & Infinium 450K) vs. reads coverage for



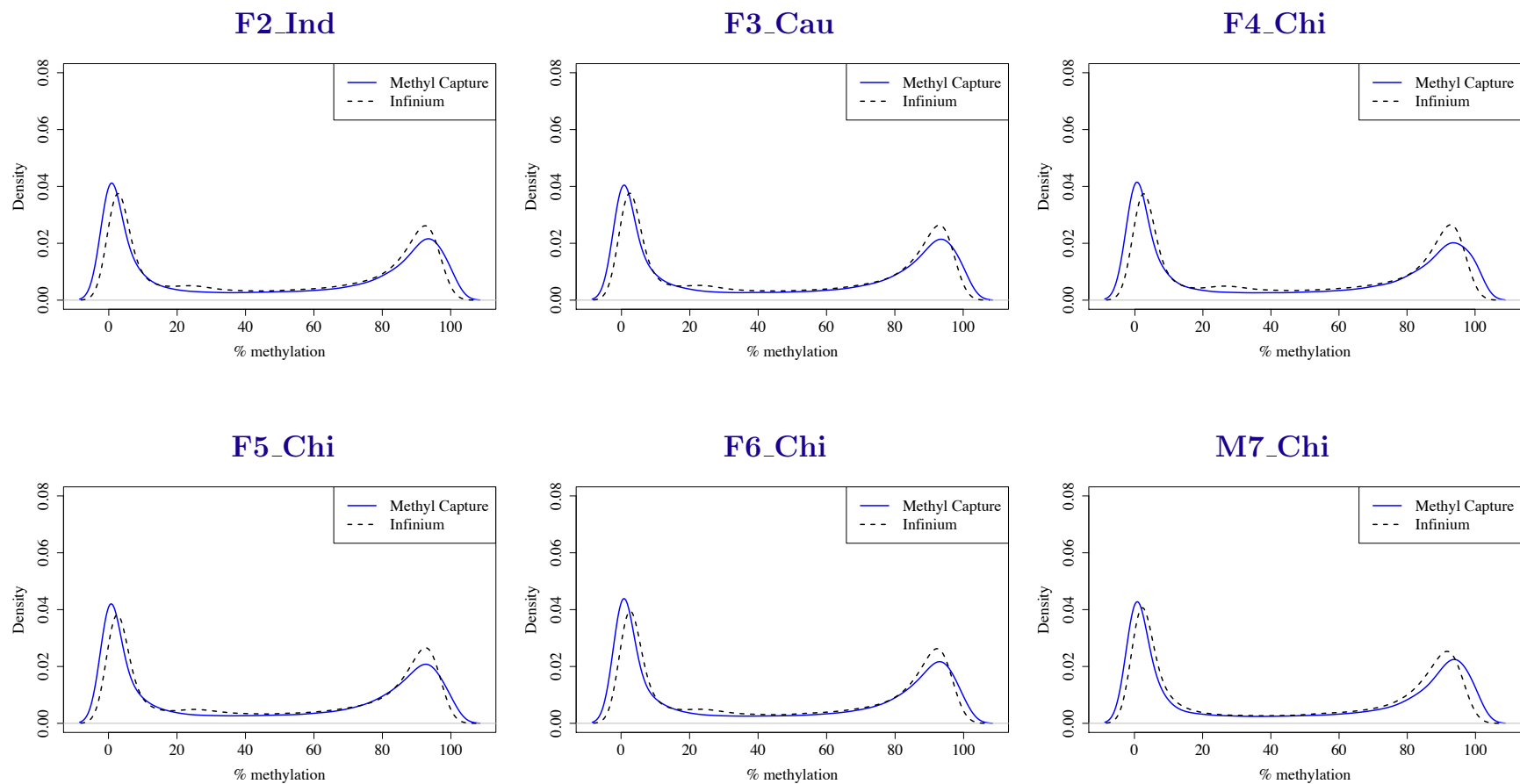
Supplementary Figure 13: Distribution of methylation values for all CpGs from MC Seq ($\geq 10X$, solid line) and Infinium 450K (dotted line).

Distribution of methylation values from MC Seq ($\geq 10X$) & Infinium 450K using all CpGs for



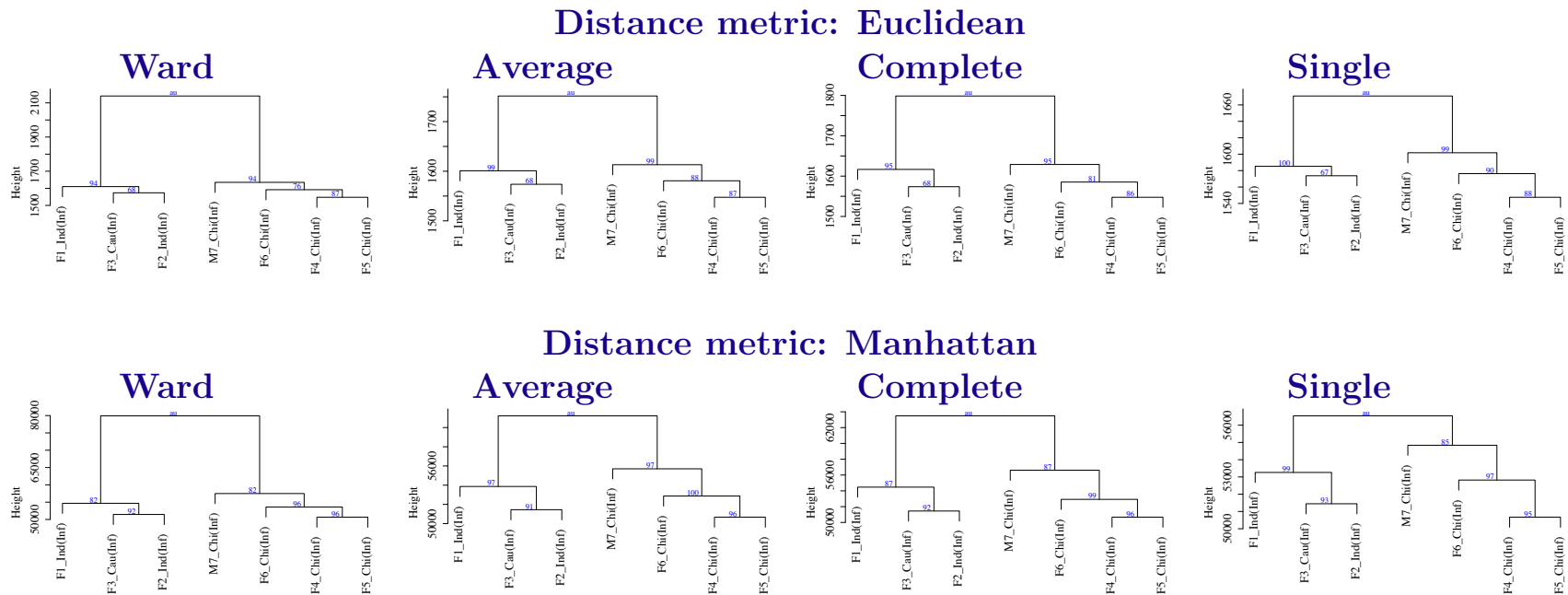
Supplementary Figure 14: Distribution of methylation values for common CpGs from MC Seq ($\geq 10X$, solid line) and Infinium 450K (dotted line).

Distribution of methylation values from MC Seq ($\geq 10X$) & Infinium 450K using common CpGs for



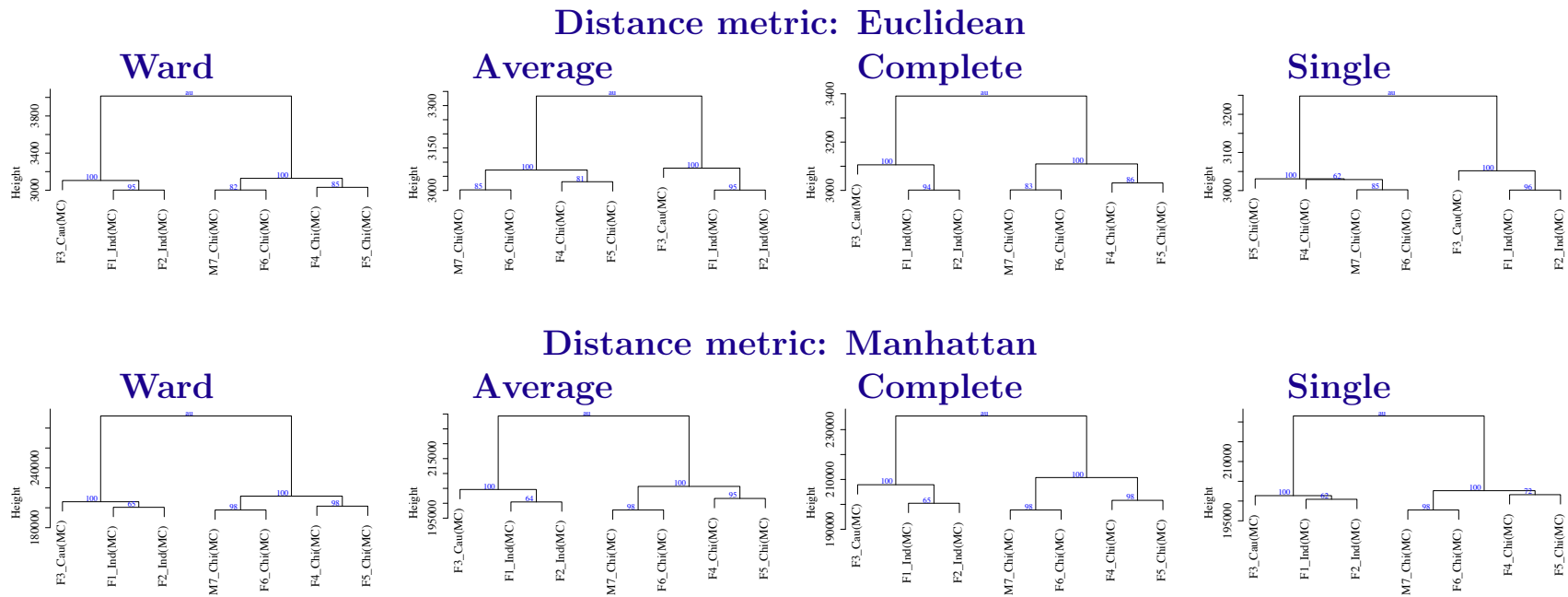
Supplementary Figure 15: Hierarchical clustering analysis of most variable CpG sites from Infinium 450K, using different distance metrics (euclidean, manhattan) and agglomeration methods (ward, average, complete, single), together with approximately unbiased (AU) p-value. Clusters with AU > 95% are considered strongly supported by data.

Hierarchical clustering analysis using Infinium 450K



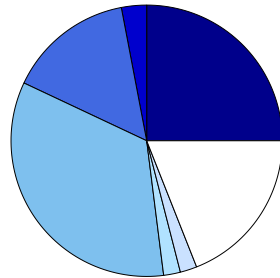
Supplementary Figure 16: Hierarchical clustering analysis of most variable CpG sites from MC Seq, using different distance metrics (euclidean, manhattan) and agglomeration methods (ward, average, complete, single), together with approximately unbiased (AU) p-value. Clusters with AU > 95% are considered strongly supported by data.

Hierarchical clustering analysis using MC Seq

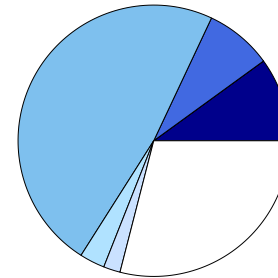


Supplementary Figure 17: Functional genomic distribution (promoter, 5'-UTR, exon, intron, 3'-UTR, TTS and intergenic) of most variable CpGs (right) vs. all CpGs (left).

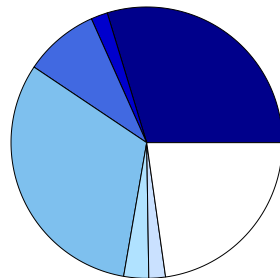
MC Seq (max)



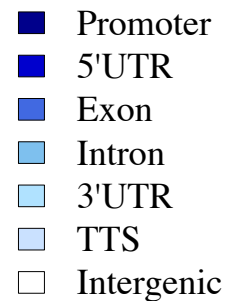
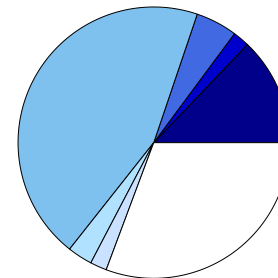
MC Seq (most variable)



Infinium 450K

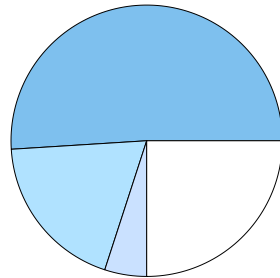


Infinium 450K (most variable)

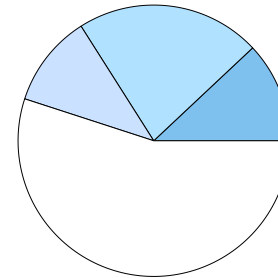


Supplementary Figure 18: CpG content distribution (island, shore, shelf and open sea) of most variable CpGs (right) vs. all CpGs (left).

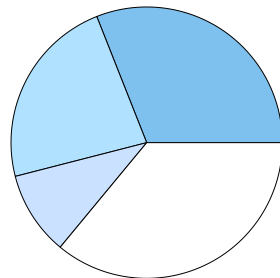
MC Seq (max)



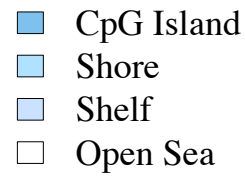
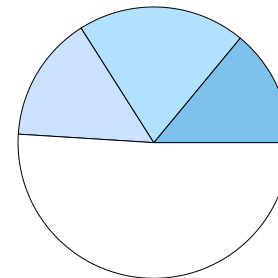
MC Seq (most variable)



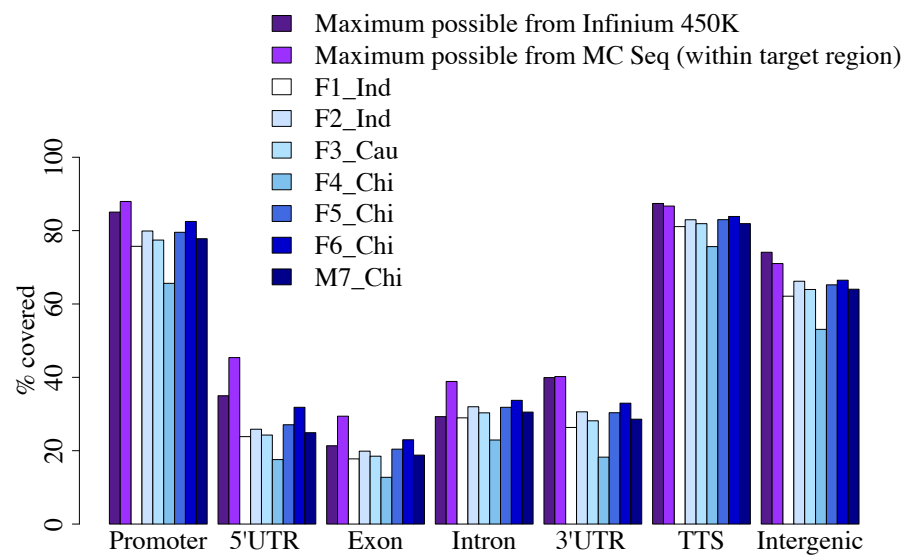
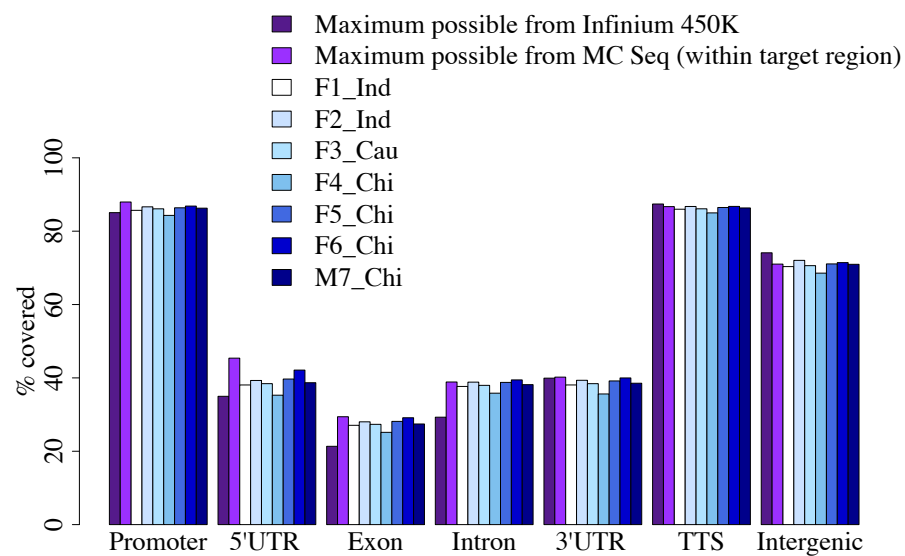
Infinium 450K



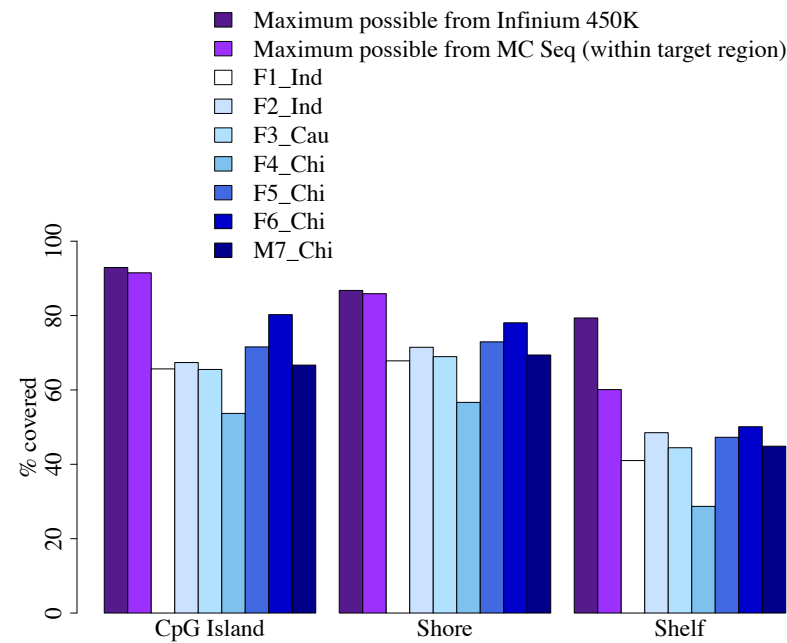
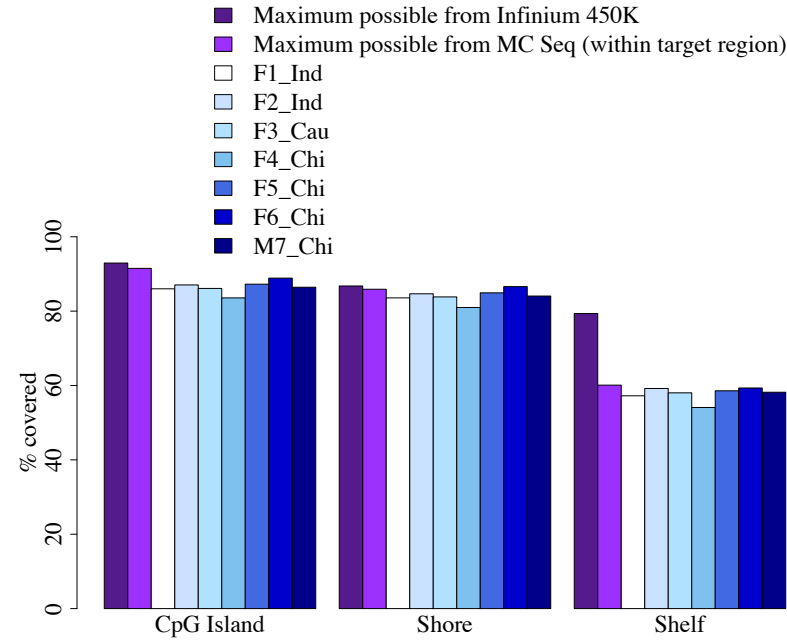
Infinium 450K (most variable)



Supplementary Figure 19: Genomic coverage at $\geq 10X$ (top) and $\geq 30X$ (bottom).

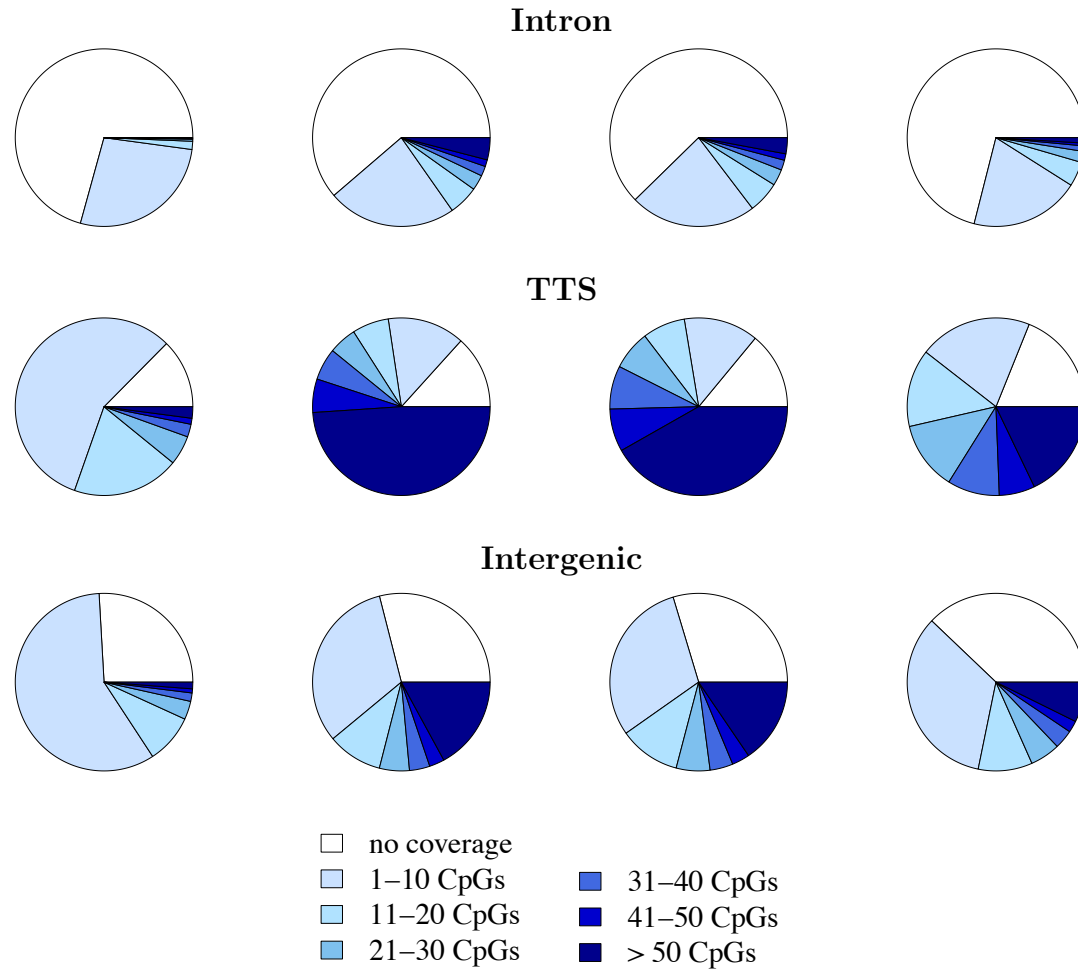


Supplementary Figure 20: CpG coverage at $\geq 10X$ (top) and $\geq 30X$ (bottom).



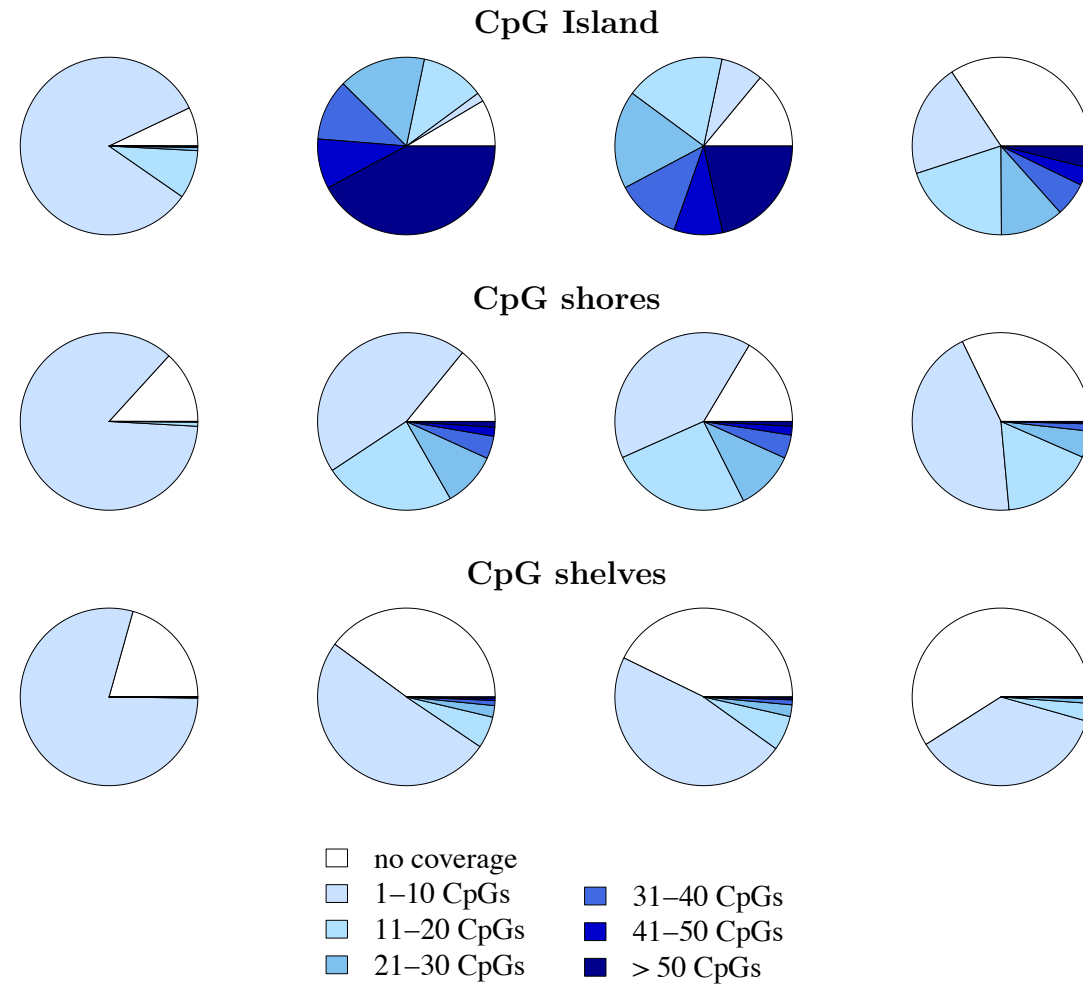
Supplementary Figure 21: Density of genomic coverage.

Infinium 450K MC Seq (max) MC Seq ($\geq 10X$) MC Seq ($\geq 30X$)



Supplementary Figure 22: Density of CpG coverage.

Infinium 450K MC Seq (max) MC Seq ($\geq 10X$) MC Seq ($\geq 30X$)



Supplementary Table 1: Summary of sequence alignment and duplicate rates for four technical replicates.

	F1_Ind	F3_Cau	F4_Chi	M7_Chi
Raw sequence reads	93798516	96096291	92769553	82161548
Sequence pairs analysed in total	83536052	83171725	82577165	74770924
Number of paired-end alignments with a unique best hit	69907181	70538589	68090749	58664868
Mapping efficiency (%)	83.7	84.8	82.5	78.5
Duplicate (%)	47.48	39.53	52.6	42.68
Sequence pairs after removing duplicate	36712773	42658112	32276416	33625981
Reads in targeted region (%)	89.43	88.81	89.56	90.82
Reads in targeted regions 200 bp (%)	92.91	92.32	92.84	93.35

Supplementary Table 2. Summary of sequence alignment and duplicate rates for two buccal epithelium samples using 1µg DNA from MC Seq.

	F5_Chi	M7_Chi	Average
Raw sequence reads	82669270	46087480	64378375
Sequence pairs analysed in total	73443192	41243404	45866620
Number of paired-end alignments with a unique best hit	68026518	37965457	42481376
Mapping efficiency (%)	92.6	92.1	92.77
Duplicate (%)	29.66	12.44	16.73
Sequence pairs after removing duplicate	47848930	33244243	33603102
Reads in targeted region (%)	93.07	93.4	93.42
Reads in targeted regions 200 bp (%)	96.77	96.81	96.85

Supplementary Table 3. Summary of technical replicates performance for Infinium 450k.

Sample ID	Rep1_Count	Rep2_Count	Overlapped CpGs	Pearson R
F2_Ind	431769	431755	431581	0.9988
F5_Chi	431791	431779	431628	0.9988
F6_Chi	431779	431755	431599	0.9987
M7_Chi	431798	431772	431631	0.9988